

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/235716063>

# Superpixel Classification Based Optic Disc and Optic Cup Segmentation for Glaucoma Screening

Article · February 2013

DOI: 10.1109/TMI.2013.2247770 · Source: PubMed

CITATIONS

249

READS

801

10 authors, including:



**Jun Cheng**

Institute for Infocomm Research

70 PUBLICATIONS 838 CITATIONS

[SEE PROFILE](#)



**Yanwu Xu**

Institute for Infocomm Research

88 PUBLICATIONS 1,210 CITATIONS

[SEE PROFILE](#)



**Ching-yu Cheng**

Duke-NUS Medical School

541 PUBLICATIONS 13,502 CITATIONS

[SEE PROFILE](#)



**Tin Aung**

National University of Singapore

726 PUBLICATIONS 22,541 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Diabetic Macular Edema Study [View project](#)



Singapore epidemiology of eye disease study [View project](#)

# Supervoxel Classification based Optic Disc and Optic Cup Segmentation for Glaucoma Screening

Jun Cheng\*, Jiang Liu, Yanwu Xu, Fengshou Yin, Damon Wing Kee Wong, Ngan-Meng Tan, Dacheng Tao, Ching-Yu Cheng, Tin Aung and Tien Yin Wong

**Abstract**—Glaucoma is a chronic eye disease that leads to vision loss. As it cannot be cured, detecting the disease in time is important. Current tests using intraocular pressure (IOP) are not sensitive enough for population based glaucoma screening. Optic nerve head assessment in retinal fundus images is both more promising and superior. This paper proposes optic disc and optic cup segmentation using supervoxel classification for glaucoma screening. In optic disc segmentation, histograms and centre surround statistics are used to classify each supervoxel as disc or non-disc. A self-assessment reliability score is computed to evaluate the quality of the automated optic disc segmentation. For optic cup segmentation, in addition to the histograms and centre surround statistics, the location information is also included into the feature space to boost the performance. The proposed segmentation methods have been evaluated in a database of 650 images with optic disc and optic cup boundaries manually marked by trained professionals. Experimental results show an average overlapping error of 9.5% and 24.1% in optic disc and optic cup segmentation, respectively. The results also show an increase in overlapping error as the reliability score is reduced, which justifies the effectiveness of the self-assessment. The segmented optic disc and optic cup are then used to compute the cup to disc ratio for glaucoma screening. Our proposed method achieves areas under curve of 0.800 and 0.822 in two data sets, which is higher than other methods. The methods can be used for segmentation and glaucoma screening. The self-assessment will be used as an indicator of cases with large errors and enhance the clinical deployment of the automatic segmentation and screening.

**Index Terms**- Optic disc segmentation, optic cup segmentation, glaucoma screening

## I. INTRODUCTION

Glaucoma is a chronic eye disease in which the optic nerve is progressively damaged. It is the second leading cause of

blindness, and is predicted to affect around 80 million people by 2020 [1]. Progression of the disease leads to loss of vision, which occurs gradually over a long period of time. As the symptoms only occur when the disease is quite advanced, glaucoma is called the silent thief of sight. **Glaucoma cannot be cured, but its progression can be slowed down by treatment.**

Therefore, detecting glaucoma in time is critical. However, many glaucoma patients are unaware of the disease until it has reached its advanced stage. In Singapore, more than 90% of patients are unaware that they have this condition [2][3]. In Australia, about 50% of people with glaucoma are undiagnosed [4]. Since glaucoma progresses with few signs or symptoms and the vision loss from glaucoma is irreversible, screening of people at high risk for the disease is vital. There are three methods to detect glaucoma: (1) assessment of raised intraocular pressure (IOP), (2) assessment of abnormal visual field, (3) assessment of damaged optic nerve head. The IOP measurement using non-contact tonometry (also known as the “airpuff test”) is neither specific nor sensitive enough to be an effective screening tool because glaucoma can be present with or without increased IOP. A functional test through vision loss requires special equipments only present in territory hospitals and therefore unsuitable for screening. Assessment of the damaged optic nerve head is both more promising, and superior to IOP measurement or visual field testing for glaucoma screening. Optic nerve head assessment can be done by a trained professional. However, manual assessment is subjective, time consuming and expensive. Therefore, automatic optic nerve head assessment would be very beneficial.

One strategy for automatic optic nerve head assessment is to use image features for a binary classification between glaucomatous and healthy subjects [5][6][7]. These features are normally computed at the image-level. In these methods, selection of features and classification strategy is difficult and challenging [8]. The other strategy is to follow the clinical indicators. Many glaucoma risk factors are considered, such as the vertical cup to disc ratio (CDR)[9], disc diameter [10], ISNT [11] rule, peripapillary atrophy (PPA) [12], notching [13], etc. Although different ophthalmologists have different opinions on the usefulness of these factors, CDR is well accepted and commonly used. A larger CDR indicates a higher risk of glaucoma. There has been some research into automatic CDR measurement from 3D images [14][15][16][17]. However, because 3D images are not easily available, 2D colour fundus images are still referred to by most clinicians [8][14]. Moreover, the high cost of obtaining 3D images make it inappropriate for a large-scale screening program [8]. This paper focuses on automatic glaucoma screening using CDR

This work was supported in part by the Agency for Science, Technology and Research, Singapore, under SERC grant 121-148-0007. Asterisk indicates corresponding author.

\*J. Cheng is with iMED Ocular Imaging Programme in Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore (email: jcheng@i2r.a-star.edu.sg).

J. Liu, Y. Xu, F. Yin, D. W. K. Wong, N. M. Tan are with iMED Ocular Imaging Programme in Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore (email: jliu@i2r.a-star.edu.sg, yaxu@i2r.a-star.edu.sg, fyin@i2r.a-star.edu.sg, wkwong@i2r.a-star.edu.sg, nmtan@i2r.a-star.edu.sg).

D. Tao is with the Centre for Quantum Computation & Intelligent Systems and the Faculty of Engineering & Information Technology, University of Technology, Sydney, 235 Jones Street, Ultimo, NSW 2007, Australia (Email: dacheng.tao@uts.edu.au).

C. Y. Cheng is with the Department of Ophthalmology at the National University of Singapore (email: ching-yu\_cheng@nuhs.edu.sg).

T. Aung and T. Y. Wong are with the Singapore Eye Research Institute, Singapore National Eye Centre and the Department of Ophthalmology at the National University of Singapore (email: aung.tin@sneec.com.sg, tien\_yin\_wong@nuhs.edu.sg).

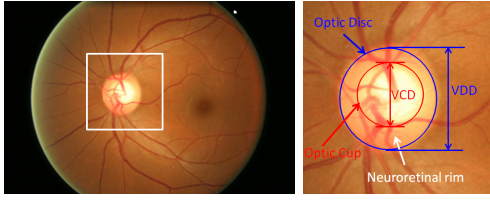


Fig. 1. Major structures of the optic disc: The region enclosed by the blue line is the optic disc; the central bright zone enclosed by the red line is the optic cup; and the region between the red and blue lines is the neuroretinal rim.

from 2D fundus images.

The optic nerve head or the optic disc (in short, disc) is the location where ganglion cell axons exit the eye to form the optic nerve, through which visual information of the photo-receptors is transmitted to the brain. In 2D images, the disc can be divided into two distinct zones; namely, a central bright zone called the optic cup (in short, cup) and a peripheral region called the neuroretinal rim. Fig. 1 shows the major structures of the disc. The CDR is computed as the ratio of the vertical cup diameter (VCD) to vertical disc diameter (VDD) clinically. Accurate segmentations of disc and cup are essential for CDR measurement. Several methods [8][18][19][20][21] have been proposed for automatic CDR measurement from 2D fundus images.

This paper proposes superpixel classification based disc and cup segmentations for glaucoma screening. A similar concept has been used for vessel segmentation [22]. We compute centre surround statistics from superpixels and unify them with histograms for disc and cup segmentation. We incorporate prior knowledge of the cup by including location information for cup segmentation. Based on the segmented disc and cup, CDR is computed for glaucoma screening. In addition, the proposed method computes a self-assessment reliability score for its disc segmentation result. Self-assessment is an important issue that has previously seldom been discussed in disc segmentation. In practice, an automated segmentation method might work well for most images while working poorly for the rest. Therefore, it is important to have self-assessment where users are warned of cases with potentially large errors.

The paper is organized as follows. In Section II, we introduce superpixel classification based OD segmentation including the generation of superpixels, the extraction of features from superpixels for the classification and the computation of the self-assessment reliability score. Section III introduces superpixel classification based cup segmentation, where the procedure is similar to that in disc segmentation. Section IV shows the experimental results including the accuracy for disc and cup segmentation as well as glaucoma screening. Discussions and conclusions are presented in final section.

## II. OPTIC DISC SEGMENTATION

### A. Background

Localization and segmentation of disc are very important in many computer aided diagnosis systems, including glaucoma screening. The localization focuses on finding an disc pixel, very often the centre. It has been extensively studied for

applications in diabetic screening [23][24]. Our work focuses on the segmentation problem and the disc is located by our earlier method in [25], which works well in our data set for glaucoma screening as there are few white lesions to confuse disc localization as compared to diabetic screening. The segmentation estimates the disc boundary, which is a challenging task due to blood vessel occlusions, pathological changes around disc, variable imaging conditions, etc.

Some approaches have been proposed for disc segmentation, which can be generally classified as template based methods [26][27][28], deformable model based methods [29][14][30][8][31][32] and pixel classification based methods [15][33]. In [26][27], circular Hough transform is used to model the disc boundary because of its computational efficiency. However, clinical studies have shown that a disc has a slightly oval shape with the vertical diameter being about 7%-10% larger than the horizontal one [34]. Circular fitting might lead to an under-estimated disc and an over-estimated CDR, so ellipse fitting is often adopted for glaucoma detection [28]. In [29], Lowell *et al.* employed the active contour model, which consists of finding the optimal points based on the image gradient and the smoothness of the contour. In [14], Xu *et al.* employed the deformable model technique through minimization of the energy function defined by image intensity, image gradient, and boundary smoothness. In [30], a level set is used to estimate the disc followed by an ellipse fitting to smooth the boundary. In [8], the authors proposed a modified Chan-Vese model using texture features. In [31], edge detection and circular Hough transform are combined with an active shape model to extract the disc. Recently, the active shape model is also applied on probability maps to find the disc boundary [32]. In addition, we also present a superpixel classification based approach using histograms [35] to improve the initialization of the disc for deformable models. Both the template and deformable model based methods are based on the edge characteristics. The performance of these methods very much depends on the differentiation of edges from the disc and other structures, especially the PPA. Examples of PPA are shown as the area between the blue and green lines in Fig. 2. The PPA region is often confused as part of disc for two reasons: 1) it looks similar to the disc; 2) its crescent shape [36] makes it form another ellipse (often stronger) together with the disc. Deformable models are sensitive to poor initialization. Very often, the deformation cannot exclude PPA from the segmented disc if it has been included in the initialization. For example, the red line in the first example in Fig. 2 is the boundary detected by the active shape model based method in [31]. To overcome the problem, we previously proposed a template based approach with PPA elimination [28]. By using a PPA detection module based on texture, our method reduces the chance of mistaking PPA as part of the disc. However, the approach does not work well when the PPA area is small, or when the texture is not significantly predominant such as the second example in Fig. 2. Pixel classification based methods use various features such as intensity, texture, etc. from each pixel and its surroundings to find the disc. Muramatsu *et al.* [37] compared the pixel classification based methods with the deformable model based

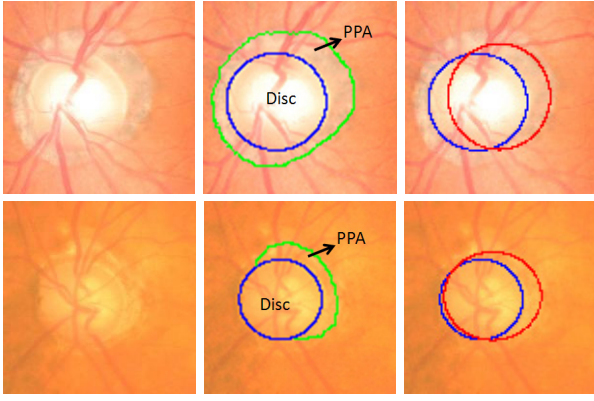


Fig. 2. Challenge in disc segmentation. Blue lines: manual disc boundary; red lines: automated disc boundary by [31] and [28] in the first and second example respectively; green lines: the PPA boundary.

methods and concluded that their performances were similar. Moreover, the number of pixels is high even at moderate resolutions, which makes the optimization on the level of pixels intractable [38].

To overcome the limitations of pixel classification based methods and deformable model based methods, we propose a superpixel classification based method and combine it with the deformable model based methods. Superpixels are local, coherent and provide a convenient primitive to compute local image features. They capture redundancy in the image and reduce the complexity of subsequent processing. In the proposed method, superpixel classification is used for an initialization of disc boundary and the deformable model is used to fine tune the disc boundary, i.e., a superpixel classification based disc initialization for deformable models. The flow chart of the proposed disc segmentation method is summarized in Fig. 3. The segmentation comprises: a superpixel generation step to divide the image into superpixels; a feature extraction step to compute features from each superpixel; a classification step to determine each superpixel as a disc or non-disc superpixel to estimate the boundary; a deformation step using deformable models to fine tune the disc boundary.

### B. Superpixel Generation

Many algorithms have been proposed for superpixel classification [39][40][41][42][43][44][45][46][47]. They have been proved to be useful in image segmentations in various images of scene, animal, human etc. This paper uses the simple linear iterative clustering [47] algorithm (SLIC) to aggregate nearby pixels into superpixels in retinal fundus images. Compared with other superpixel methods, SLIC is fast, memory efficient and has excellent boundary adherence. SLIC is also simple to use with only one parameter, i.e., the number of desired superpixels  $k$ . Here we give a brief introduction of the SLIC algorithm while more details of the algorithms can be found in the SLIC paper [47].

In SLIC,  $k$  initial cluster centres  $C_k$  are sampled on a regular grid spaced by  $S = \sqrt{N/k}$  pixels apart from the image with  $N$  pixels. The centres are first moved towards the lowest gradient position in a  $3 \times 3$  neighborhood. Clustering is then applied.

For each  $C_k$ , SLIC iteratively searches for its best matching pixel from the  $2S \times 2S$  neighborhood around  $C_k$  based on color and spatial proximity and then compute the new cluster centre based on the found pixel. The iteration continues until the distance between the new centres and previous ones is small enough. Finally, a post-processing is applied to enforce connectivity.

### C. Feature Extraction

1) *Contrast Enhanced Histogram*: Many features such as colour, appearance, gist, location and texture can be extracted from superpixels for classification [48]. Since colour is one of the main differences between disc and non-disc region, colour histogram from superpixels is an intuitive choice [35]. Motivated by the large contrast variation between images and the use of histogram equalization in biological neural networks [49], histogram equalization is applied to red  $r$ , green  $g$ , and blue  $b$  channels from RGB colour spaces individually to enhance the contrast for easier analysis. However, histogram equalization on  $r, g, b$  may yield dramatic changes in the image's colour balance. Thus, hue  $h$  and saturation  $s$  from HSV colour space are also included to form five channel maps. The value  $v$  from HSV colour space is not used as it is almost the same as the red channel for retinal images. The histogram of each superpixel is computed from all the five channels: the histogram equalized  $r, g, b$  as well as the original  $h, s$ . The histogram computation uses 256 bins and  $256 \times 5 = 1280$  dimensional feature  $HIST_j = [\hat{h}_j(HE(r)) \ \hat{h}_j(HE(g)) \ \hat{h}_j(HE(b)) \ \hat{h}_j(h) \ \hat{h}_j(s)]$  is computed for the  $j^{th}$  superpixel  $SP_j$ , where  $HE(\cdot)$  denotes the function of histogram equalization and  $\hat{h}_j(\cdot)$  the function to compute histogram from  $SP_j$ .

2) *Centre surround statistics*: As we described earlier, the PPA region looks to be close to the disc. It is important to include features that reflect the difference between the PPA region and the disc region. The superpixels from the two regions often appear similar except for the texture: the PPA region contains blob-like structures while the disc region is relatively more homogeneous. The histogram of each superpixel does not work well as the texture variation in the PPA region is often from a larger area than the superpixel [35]. This is because the superpixel often consists of a group of pixels with similar colours. Inspired by these observations, we propose centre surround statistics (CSS) from superpixels as a texture feature.

To compute CSS, nine spatial scale dyadic Gaussian pyramids are generated with a ratio from 1:1 (level 0) to 1:256 (level 8) as illustrated in Fig. 4. Multiple scales are used as the scale of the blob-like structures largely vary. The dyadic Gaussian pyramid [50] is a hierarchy of low-pass filtered versions of an image channel, so that successive levels correspond to lower frequencies. It is accomplished by convolution with a linearly separable Gaussian filter and decimation by a factor of two. Then centre surround operation between centre (finer) levels  $c = 2, 3, 4$  and surround levels (coarser)  $s = c + d$ , with  $d = 3, 4$  is applied to obtain six maps empirically computed at levels of 2-5, 2-6, 3-6, 3-7, 4-7, and 4-8 from an image channel



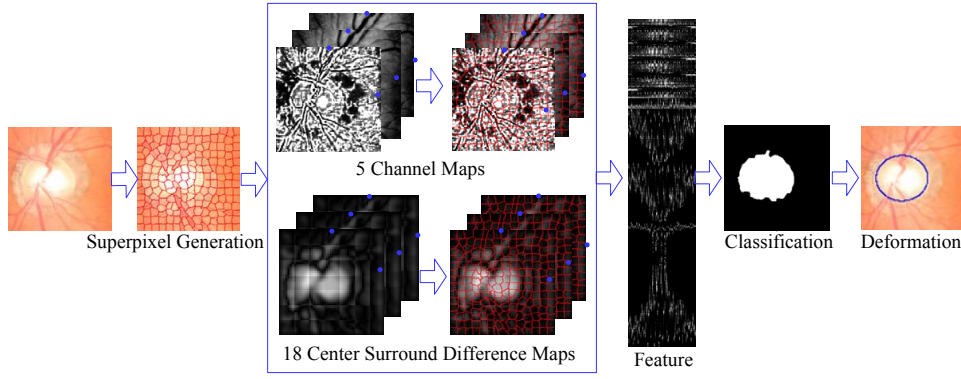


Fig. 3. Superpixel based Optic Disc Segmentation: Each image is divided into superpixels. The features computed from 18 centre surround difference maps and 5 channel maps are used to classify the superpixels as disc or non-disc. The  $f^h$  column in the feature map corresponds to the feature for the  $f^h$  superpixel.

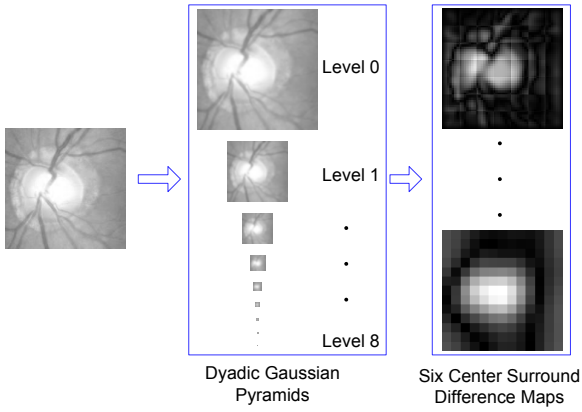


Fig. 4. Centre surround difference from an image channel.

[51][52][53]. Denote the feature map in centre level  $c$  as  $I(c)$  and the feature map in surround level  $s$  as  $I(s)$ . Because of the scale difference, we first interpolate  $I(s)$  to be the same size as  $I(c)$  and the interpolated map is denoted as  $f_{s-c}(I(s))$ , where  $f_{s-c}(\cdot)$  denotes the interpolation from the surround level  $s$  to the centre level  $c$ . The centre surround difference is then computed as  $|I(c) - f_{s-c}(I(s))|$ . All the difference maps are resized to be the same size as the original. We compute the maps from  $r$ ,  $g$  and  $b$  channels to get  $6 \times 3 = 18$  maps. The CSS features are then computed as the first and second moments of these maps within superpixels. Denoting  $M_i$ ,  $i = 1, 2, \dots, 18$ , as the  $i^{th}$  map, the feature  $CSS_j$  consists of the mean  $\mu_j$  and variance  $var_j$  of maps within the superpixels, i.e.,  $CSS_j = [\mu_j \text{ } var_j]$ , where  $\mu_j$  and  $var_j$  from superpixel  $SP_j$  with  $n_j$  pixels are computed by:

$$\mu_j(i) = \frac{1}{n_j} \sum_{(x,y) \in SP_j} M_i(x,y) \quad (1)$$

$$var_j(i) = \frac{1}{n_j} \sum_{(x,y) \in SP_j} (M_i(x,y) - \mu_j(i))^2 \quad (2)$$

3) *Final Feature*: Since the texture feature from the PPA region is often involved in a large region, the features from neighboring superpixels are also considered in the classification of the current superpixel. We empirically search for

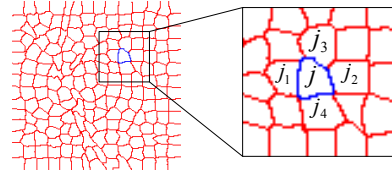


Fig. 5. Illustration of neighboring superpixels

four neighboring superpixels for  $SP_j$  and denote them as  $SP_{j_1}$ ,  $SP_{j_2}$ ,  $SP_{j_3}$  and  $SP_{j_4}$ .  $SP_{j_1}$  is determined as the first superpixel by moving out of the current superpixel horizontally to the left from its centre. Similarly,  $SP_{j_2}$ ,  $SP_{j_3}$  and  $SP_{j_4}$  are determined by moving right, up and down, as illustrated in Fig. 5. The CSS feature for  $SP_j$  would then be expanded to:  $\widehat{CSS}_j = [CSS_j \text{ } CSS_{j_1} \text{ } CSS_{j_2} \text{ } CSS_{j_3} \text{ } CSS_{j_4}]$ , which has a dimension of  $18 \times 2 \times 5 = 180$ . We combine  $HIST_j$  and  $\widehat{CSS}_j$  to form the proposed feature  $[HIST_j \text{ } \widehat{CSS}_j]$ . To avoid large scale features dominating those in a smaller scale, they are each normalized to the range of [0 - 1], with  $\ell_1$  - normalization of each histogram.

#### D. Initialization and Deformation

A support vector machine is used as the classifier. The LIBSVM [54] with linear kernel is used in our experiments. In this paper, the linear kernel is used instead of non-linear radial basis function (RBF) kernel as the feature dimensionality is high. We find that the nonlinear mapping using the RBF kernel does not improve the performance. In the training, we randomly obtain the same number of superpixels from the disc and non-disc region from a set of training images with manual disc boundary. One challenge to find a good classifier is that samples from the non-disc region are often from different clusters with unbalanced numbers. One typical example is PPA. There are often fewer superpixels from the PPA region compared with other non-disc region, and the trained classifier is often dominated by superpixels from the latter. To overcome the problem, we adopt a bootstrapping strategy [55]: we first restrict the active training data set to be a sub-set of the available training data set (pool) and proceed iteratively. In every iteration, the training is performed on the

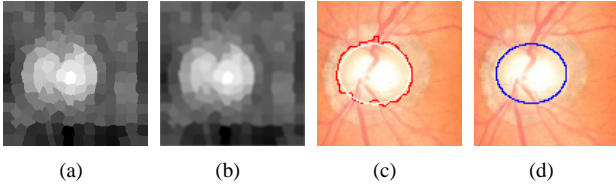


Fig. 6. Illustration of deformation tuning: (a) decision values (b) smoothed decision values (c) the raw estimation (red) and the fitted estimation (white)(d) final result after deformation.

active set and returns a preliminary classifier. The classifier is then used to evaluate the pool. In every round of training, the active training set is extended by examples misclassified in the previous round, thus emphasizing samples close to the decision boundary. We repeat the iterations until there is no improvement in the classification accuracy or the maximum iterations have been reached. After that, the trained classifier is used for subsequent testing.

Instead of directly using the binary classification results from LIBSVM, the output values from the SVM decision function are used. The output value for each superpixel is used as the decision values for all pixels in the superpixel as shown in Fig. 6(a). A smoothing filter is then applied on the decision values to achieve smoothed decision values, as in Fig. 6(b). In our implementation, mean filter and gaussian filter are tested and the mean filter is found to be a better choice. The smoothed decision values are then used to obtain the binary decisions for all pixels with a threshold. In our experiments, we assign +1 and -1 to positive (disc) and negative (non-disc) samples, and the threshold is the average of them, i.e., 0. After we get the binary decisions for all pixels, we have a matrix with binary values with 1 as object and 0 as background. The largest connected object, i.e., the connected component with largest number of pixels, is obtained through morphological operation and its boundary is used as the raw estimation of the disc boundary. The best fitted ellipse using elliptical Hough transform [28] is computed as the fitted estimation. The active shape model employed in [31] is used to fine tune the disc boundary. Compared with [31], the proposed method can also be treated as an active shape model based approach with initial contour obtained by superpixel classification.

#### E. Self-assessment reliability score

As the disc is often close to an ellipse, the obtained boundary before and after ellipse fitting should be close if the superpixel based segmentation is close to the actual boundary. Otherwise, the result is likely to be less reliable. Inspired by this, **we propose to compute a self-assessment reliability score for the segmentation.** Define the set of points from the raw estimation as  $X$  and the set of points from the fitted estimation as  $Y = f(X)$ , e.g., the red and white lines in Fig. 6(c), respectively. For each point  $x$  in  $X$ , we find its nearest point in  $Y$  and their distance is computed as

$$d_f(x) = \inf\{d(x, y) | y \in Y\} \quad (3)$$

where  $\inf$  represents the infimum and  $d(x, y)$  the Euclidean distance between  $x$  and  $y$ . Then, the self-assessment reliability

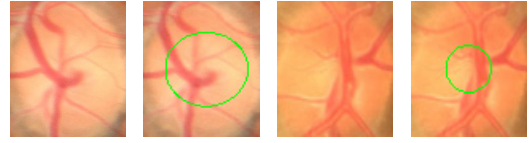


Fig. 7. Samples of discs without obvious pallor: green lines are manual cup boundary

score is computed as the ratio of the number of  $x$  with  $d_f(x) < T$  to the total number of  $x$ , i.e.,

$$r(X) = \frac{Card(\{x | d_f(x) < T, x \in X\})}{Card(X)}, \quad (4)$$

where  $Card(Z)$  is the cardinality of the set  $Z$ , and  $T$  is a threshold empirically set as five pixels in this paper for our images with average disc diameter around 350 pixels.

### III. OPTIC CUP SEGMENTATION

Detecting the cup boundary from 2D fundus images without depth information is a challenging task as depth is the primary indicator for the cup boundary. In 2D fundus images, one landmark to determine the cup region is the pallor, defined as the area of maximum colour contrast inside the disc [14]. Another landmark is the vessel bends at the boundary of the cup [13][34]. Compared with disc segmentation, fewer methods have been proposed for cup segmentation from 2D fundus images. Thresholding is used to determine the cup in [18][19][56], relying on intensity difference between cup and neuroretinal rim. A level set based approach is used in [20]. It relies on the edges between cup and neuroretinal rim. This method and thresholding based methods are essentially based on pallor information. However, in many subjects from screening data, there is no obvious pallor or edge within the disc to mark the cup boundary. Fig. 7 shows two examples of such discs. The contrast between the cup and the neuroretinal rim in the two examples are much weaker than that in Fig. 1. In [57], small vessel bends (“kinks”) from the vicinity of the initial estimated cup have been used to aid the cup segmentation. The challenge is to exclude vessel bends from a non-cup boundary, especially when the initial estimation is inaccurate. A similar concept is used in [8] to locate relevant-vessel bends (“r-bend”) at the vicinity of a pallor region determined by bright pixels. This method, again, requires pallor information to find a good initial estimation of the cup boundary. Moreover, it requires at least a few bends in nasal, inferior and superior angular of the disc for the cup boundary fitting, which is not necessarily true for many images from our experience. Xu *et al.* proposed a sliding window and regression based method [58]. Although it performs better than earlier methods, the sliding window strategy requires heavy computational cost. Recently, Yin *et al.* [59] developed a deformable model based method for cup segmentation, where the initialization of the cup boundary is based on pallor combined with prior knowledge of the cup.

The main challenge in cup segmentation is to determine the cup boundary when the pallor is non-obvious or weak. In such scenarios, we lack landmarks, such as intensity changes or edges to estimate the cup boundary reliably. Although vessel

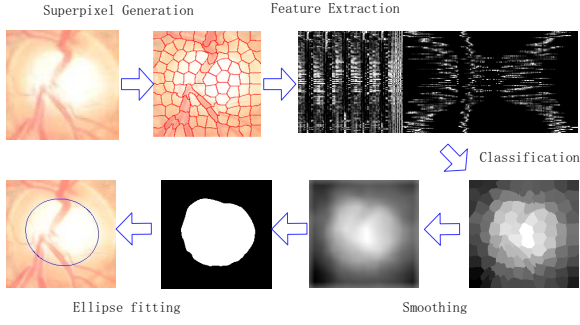


Fig. 8. Superpixel based optic cup segmentation: Each disc image is divided into superpixels. The features are used to classify the superpixels as cup or non-cup. The decision values from SVM output are smoothed to determine cup boundary.

bends are potential landmarks, they can occur at many places within the disc region and only one sub-set of these points defines the cup boundary. Besides the challenges to obtain these points, it is also difficult to differentiate the vessel bends that mark the cup boundary from other vessel bends without obvious pallor information. Moreover, combining the vessel bends with pallor information is a challenging task that often requires a set of heuristic parameters, which raises the concern of the robustness of the method. We present a superpixel classification based method for cup segmentation that incorporates prior knowledge into the training of superpixel classification instead of relying on vessel bends. Fig. 8 shows the procedure for cup segmentation, which is similar to that for disc segmentation with some minor modifications.

#### A. Feature Extraction

The feature extraction process is summarized below. After obtaining the disc, the minimum bounding box of the disc is used for cup segmentation. The histogram feature is computed similarly to that for disc segmentation, except that the histogram from the red channel is no longer used. This is because there is little information about the cup in the red channel. We denote it as  $HIST_j^c$  to be differentiated from that for disc segmentation. Similarly, the centre surround statistics  $\overline{CSS}_j^c$  can be computed. To obtain a reliable result when the pallor is non-obvious, we further include the distance  $D_j$  between the centre of superpixel  $SP_j$  and the centre of the disc as location information, as illustrated in Fig. 9. To adapt the variation of disc size,  $D_j$  is normalized distance using the height and width of the disc. Mathematically,  $D_j$  is computed as:

$$D_j = \sqrt{\left(\frac{x_c - x_j}{h}\right)^2 + \left(\frac{y_c - y_j}{w}\right)^2} \quad (5)$$

where  $(x_c, y_c)$  denotes the coordinate of the disc centre,  $(x_j, y_j)$  denotes the coordinate of the centre of  $SP_j$ ,  $h$  and  $w$  denotes the height and width of the disc, respectively. The use of  $D_j$  is inspired by the prior knowledge that the cup usually lies at the centre section of the disc. Thus, the feature vector for cup segmentation is computed as  $[HIST_j^c \ \overline{CSS}_j^c \ D_j]$ .

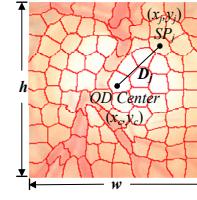


Fig. 9. Illustration of distance between centre of superpixel and centre of disc

#### B. Superpixel Classification for Optic Cup Estimation

The LIBSVM [54] with linear kernel is used again in our experiments. We randomly obtain the same number of superpixels from the cup and non-cup regions in the training step from a set of training images with manual cup boundary. Similarly, the output values from the LIBSVM decision function are used. As illustrated in Fig. 8, the output value for each superpixel is used as the decision values for all pixels in the superpixel. A mean filter is applied on the decision values to compute smoothed decision values. Then the smoothed decision values are used to obtain the binary decisions for all pixels. The largest connected object is obtained and its boundary is used as the raw estimation. The best fitted ellipse [60] is computed as the cup boundary. The ellipse fitting here is beneficial for overcoming the noise introduced by vessels especially from the inferior and superior sector of the cup. We do not apply contour deformation after we obtain the estimated cup boundary from superpixel classification, because many cases do not have an obvious/strong contrast between the cup and the rim for the deformable models. A deformation in these cases often lead to an over-estimated cup.

#### C. Cup to Disc Ratio

After obtaining the disc and cup, various features can be computed. We follow the clinical convention to compute the CDR. As mentioned in the introduction, CDR is an important indicator for glaucoma screening computed as

$$CDR = \frac{VCD}{VDD}. \quad (6)$$

The computed  $CDR$  is used for glaucoma screening. When  $CDR$  is greater than a threshold, it is glaucomatous, otherwise, healthy.

### IV. EXPERIMENTAL RESULTS

#### A. Data Sets

Our experiments uses 2326 images from 2326 different subject eyes including 650 from the Singapore Malay Eye Study (SiMES) and 1676 from the Singapore Chinese Eye Study (SCES). IOP have been measured for these eyes. Among the images, the disc and cup boundaries from 650 SiMES data set with image dimension  $3072 \times 2048$  have been manually marked by trained professionals in previous studies for disc and cup segmentation [28][58][61]. The 1676 images from SCES are collected in a screening study. There are two sizes:  $3504 \times 2336$  and  $3888 \times 2592$ . All the SCES images are resized to be the same size as the SiMES data set, for convenience.

TABLE I  
MEAN OVERLAPPING ERROR  $\mu_E$  FROM TESTING IMAGES AT DIFFERENT  
PARAMETERS FOR DISC SEGMENTATION

	no filter	$3 \times 3$	$5 \times 5$	$10 \times 10$	$15 \times 15$
100	11.0%	10.9%	10.9%	10.9%	10.9%
200	10.2%	9.6%	9.5%	9.6%	9.8%
400	10.4%	9.6%	9.6%	9.6%	10.3%

The disc localization method in [25] is then used to locate the disc and determine an  $800 \times 800$  region of interest for disc and cup segmentation. It locate disc correctly in all 650 SiMES images. In SCES, it fails in four of 1676 images. We use what the algorithm obtained for all the SCES images even if the localization fails. We evaluate the proposed disc segmentation and cup segmentation method using the manual boundary as “ground truth”. As our goal is glaucoma screening, we also compute the accuracy of the CDR computed based on the segmented disc and cup, as well as the glaucoma screening performance using the computed CDR. Among the 2326 eyes, 168 SiMES and 46 SCES eyes are diagnosed as glaucomatous by ophthalmologists. These diagnoses outcomes are used as the gold ground truth.

### B. Optic Disc Segmentation

The 650 images with manual disc and cup “ground truth” boundaries are randomly divided into 325 images for training and 325 images for testing. The overlapping error  $E$  is computed as the evaluation metric:

$$E = 1 - \frac{Area(S \cap G)}{Area(S \cup G)}, \quad (7)$$

where  $S$  and  $G$  denote the segmented and the manual “ground truth” disc respectively. We evaluate the proposed method as follows.

1) *Parameters*: There are two main parameters involved: the number of superpixels and the size of the mean filter. The first set of experiments was carried out to evaluate the performance under different parameters. Table I shows the mean overlapping error  $\mu_E$  for the number of superpixels at 100, 200 and 400 combined with different filter sizes from  $3 \times 3$  to  $15 \times 15$ , as well as the case without a filter. The number of superpixels as 200 is a good choice, neither too small to lose the precision nor too large to lose the information for classification. The size of filter, however, only affects the performance very slightly from  $3 \times 3$  to  $10 \times 10$ . The results with a filter are better than those without a filter, which justifies the effectiveness of smoothing.

2) *Effectiveness of features*: Next, we evaluate how the histogram and CSS features affect the results under the same framework in the proposed method. Table II shows the percentage of images per  $E$  interval and  $\mu_E$ . It can be seen that the unified feature performs better than a subset of the feature, which justifies the effectiveness of the combination. In particular, the comparison between the last two rows in Table II shows that combining  $\widehat{CSS}_j$  with  $HIST_j$  reduces  $\mu_E$  by  $(11.5-10.3)/11.5=10.4\%$  for images with PPA while

the improvement is minimal for images without PPA. Paired t-test shows the improvement is significant with  $p < 0.001$ . The effectiveness of  $\widehat{CSS}_j$  is justified by differentiating the PPA from the disc.

3) *Comparison with other methods*: To compare with other methods, we have conducted experiments using the previous active shape model based approach [31]. Since circular Hough transform initialization is used, it is termed “CHT-ASM”. The proposed method replaces the initialization in [31] and is termed “Proposed”. In addition, results using the modified Chan-Vese model (MCV)[8], the elliptical Hough transform (EHT) [28] and modified deformable models (MDM) [14] are also reported. Table III shows the percentage of images per  $E$  interval and  $\mu_E$  by these methods as well as the  $p$ -values in paired t-test compared with the proposed method. Compared with MCV[8], the proposed method reduces  $\mu_E$  significantly by  $(14.2-10.3)/14.2=27.5\%$  and  $(12.2-9.1)/12.2=25.4\%$  relatively for images with and without PPA, respectively. Paired t-test shows  $p < 0.001$  in both comparisons. Compared with CHT-ASM [31], the proposed method reduces  $\mu_E$  significantly by  $(13.5-10.3)/13.5=23.7\%$  with  $p = 0.003$  and  $(10.1-9.1)/10.1=9.9\%$  with  $p = 0.005$  relatively for images with and without PPA, respectively. Compared with EHT [28], the proposed method performs similarly for images without PPA, however, it reduces  $\mu_E$  significantly by a relative  $(12.2-10.3)/12.2=15.6\%$  with  $p = 0.001$  for images with PPA. Compared with MDM [14], the proposed method reduces  $\mu_E$  significantly by  $(12.4-10.3)/12.4=16.9\%$  with  $p < 0.001$  and  $(9.9-9.1)/9.9=6.6\%$  with  $p = 0.02$  relatively for images with and without PPA respectively. Moreover, the proposed disc segmentation method achieves a higher proportion of results with small errors ( $E \leq 0.2$ ) than previous methods. Fig. 10 shows six sample results by MCV, CHT-ASM, EHT, MDM and the proposed method. The first row is an example without PPA where all methods work well. The second to fifth rows are examples with PPA where the proposed method outperform most of other methods. The last row is an example where all the five methods fail to find the disc boundary accurately due to an unclear boundary between PPA and disc as well as the irregular shape of disc, however, the computed reliability score by the proposed method is 0.78, which indicates a high risk of failure. Computational cost is also evaluated. It takes 10.9 seconds per image in a dual core 3.0 GHz PC with 3.25 GB RAM, while the EHT method takes 21.4 seconds, the MDM method takes 5.2 seconds, the CHT-ASM method takes 4.5 seconds and MCV method takes 4.8 seconds.

4) *Effectiveness of the self-assessment reliability score*: To show the effectiveness of the self-assessment reliability score, we compute the overlapping errors for images at different ranges of the reliability scores. Fig. 11 shows the percentage of images in different ranges of the scores and the mean overlapping error  $\mu_E$ . An increase in overlapping error as the reliability score is reduced is shown, which demonstrates the effectiveness of the self-assessment reliability score.

5) *Performance on other data set*: Besides the local SiMES data set, we have also tested the proposed disc segmentation method on the publicly available MESSIDOR data set [62]. The MESSIDOR data set contains 1200 images with



TABLE II  
PERCENTAGE OF IMAGES PER  $E$  INTERVAL AND  $\mu_E$  USING DIFFERENT FEATURES

	$E \leq 0.05$	$E \leq 0.1$	$E \leq 0.15$	$E \leq 0.2$	$E \leq 0.25$	$\mu_E$		
						w/o PPA	with PPA	All
$CSS_j$	14%	55%	76%	84%	89%	12.3%	13.2%	12.6%
$\widehat{CSS}_j$	15%	67%	83%	92%	94%	10.2%	10.5%	10.4%
$HIST_j$ [35]	16%	65%	84%	92%	95%	9.2%	11.5%	10.0%
$HIST_j + \widehat{CSS}_j$	20%	71%	87%	93%	96%	9.1%	10.3%	9.5%

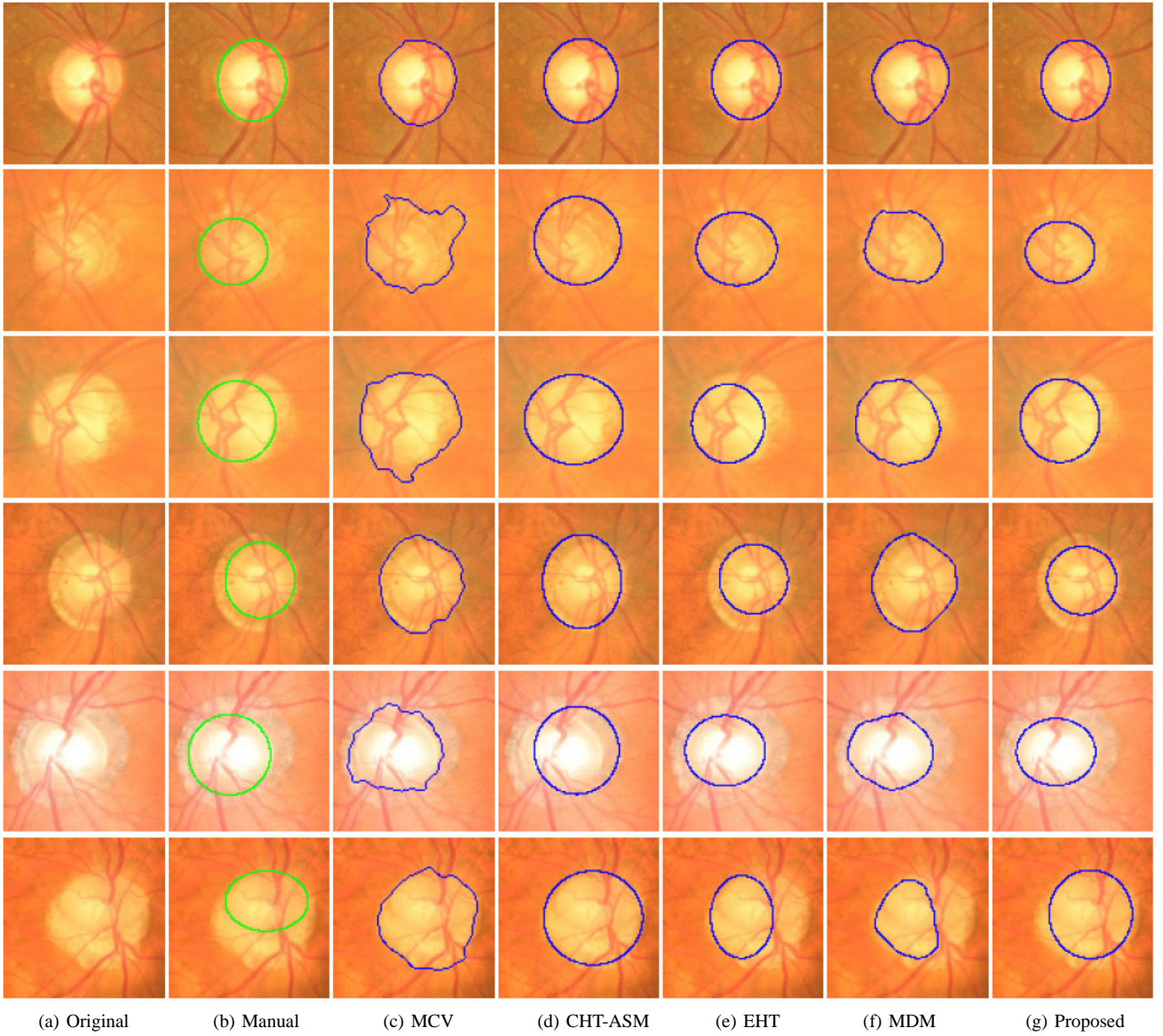
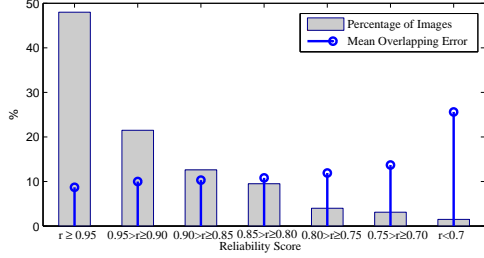


Fig. 10. Sample results. From left to right columns: (a) the original images, (b) the manual “ground truth”, (c)-(g) outlines by the MCV [8], CHT-ASM [31], EHT [28], MDM [14] and the proposed method. From top to bottom, the reliability score by the proposed method is 0.99, 0.87, 1.00, 0.97, 0.94, and 0.78 respectively. The overlapping errors by the proposed method are: 2.8%, 9.6%, 12.9%, 17.8%, 21.3%, and 37.2% respectively.

TABLE III

PERCENTAGE OF IMAGES PER  $E$  INTERVAL AND  $\mu_E$  BY VARIOUS METHODS AS WELL AS THE  $p$ -VALUES IN PAIRED T-TEST COMPARED WITH THE PROPOSED METHOD

	$E \leq 0.05$	$E \leq 0.1$	$E \leq 0.15$	$E \leq 0.2$	$E \leq 0.25$	$\mu_E$				
						w/o PPA	$p$ -value	with PPA	$p$ -value	All
MCV [8]	14%	48%	70%	82%	90%	12.2%	< 0.001	14.2%	< 0.001	12.9%
CHT-ASM	23%	64%	81%	87%	90%	10.1%	0.005	13.5%	0.003	11.3%
EHT [28]	17%	70%	85%	91%	94%	9.3%	0.455	12.2%	0.001	10.3%
MDM [14]	14%	60%	82%	91%	94%	9.9%	0.02	12.4%	< 0.001	10.8%
Proposed	20%	71%	87%	93%	96%	9.1%	-	10.3%	-	9.5%

Fig. 11. Percentage of images and mean overlapping error  $\mu_E$  per  $r$  interval

manual disc boundary available online [63]. These images have different sizes:  $1440 \times 960$ ,  $2240 \times 1488$ , or  $2304 \times 1536$  pixels. Table IV shows the comparison between our method and previous result reported in [27]. The comparison shows that our method performs better with a relative error reduction of  $(14.0 - 12.5)/14.0 = 10.7\%$  with  $p < 0.001$  in paired t-test.

### C. Optic Cup Segmentation

The same 325 training and 325 testing images from the 650 SiMES images with cup boundaries are used to evaluate the cup segmentation. In order to isolate the errors from disc segmentation, the manual discs are used in the evaluation of the cup segmentation. The overlapping error  $E$  is computed as the evaluation metric, similar to that in disc segmentation.

1) *Parameters*: The first set of experiments was carried out to evaluate the performance under different parameters. Table V shows the mean overlapping error  $\mu_E$  for the number of superpixels at 50, 100 and 200 combined with different filter sizes from  $5 \times 5$  to  $25 \times 25$  as well as the case without a filter. A number of 100 superpixels is recommended. The size of filter, however, only affects the performance very slightly from  $5 \times 5$  to  $25 \times 25$ . It is shown that the results with a filter are better than those without.

2) *Effectiveness of features*: Next, we also evaluate how the features  $\widehat{CSS}_j^c$  and  $D_j$  affect the cup segmentation. Table VI shows the percentage of images per  $E$  interval and  $\mu_E$  for cup segmentation. Fig. 12 shows the cup boundaries before ellipse fitting from five samples. The first three rows are from normal subject eyes and the last two rows are from glaucomatous eyes. The results show that  $HIST_j^c$  alone works poorly in this application. One of the reasons is that it is very sensitive to unbalanced illumination across the disc. For example, the temporal sector of the disc is often brighter than the nasal sector and the obtained results are often biased to

the temporal sector. This is also one of the reasons that a threshold based approach works poorly. Combining  $\widehat{CSS}_j^c$  with  $HIST_j^c$  reduces the error because  $\widehat{CSS}_j^c$  encodes the difference through centre surround operation. However, we still observe some bad results for the disc without obvious pallor such as the second example in Fig. 12. The location information  $D_j$  plays an important role as it is an important prior knowledge that the cup is usually the centre section of the disc. Therefore, the further away from the disc centre, the less likely the superpixel is from the cup. By including  $D_j$  into the feature space, the obtained result is closer to the manual one.

3) *Comparison with other methods*: To compare with other methods, we have also conducted experiments using the threshold method [56], r-bend [8], the ASM method [59] and the regression method [58]. Table VII shows the percentage of images per  $E$  interval and  $\mu_E$  using the manual cup as “ground truth”. The relative reductions of overlapping error by the proposed method are 55.0%, 39.0%, 23.0% and 15.1% compared with the threshold, r-bend, ASM and regression methods, respectively. Paired t-test shows  $p < 0.001$  in all the four comparisons. Moreover, the proposed cup segmentation method also achieves a higher proportion of results with smaller errors than previous methods. Fig. 13 shows the final cup boundaries for the same examples by the proposed method and four other methods. From the results, we can see that although previous methods work well when pallor is very obvious, their performance in cup segmentation without very obvious pallor is less accurate. The computational cost is also evaluated. It takes about 2.6 seconds per image in a dual core 3.0 GHz PC with 3.25 GB RAM in comparison with more than 6 minutes by the regression method and 0.85 second by ASM. A limitation of the proposed method is that it over-estimates very small cups ( $CDR < 0.4$ ) and under-estimates very large cups ( $CDR > 0.8$ ) when the pallor is not obvious, e.g., the third and last samples in Fig. 13. This is because there are very few such very small or very large cups in the data set. When  $D_j$  is included in the feature space where most discs have medium cups, the trained classifier is dominated by the majority medium cups. This limitation is reasonable as these are the most difficult cases even for human eyes. Although the bias exists, the obtained CDRs for very large cups are still high and the CDRs for very small cups are still small, as shown in the examples in Fig. 13.

TABLE IV  
PERCENTAGE OF IMAGES PER  $E$  INTERVAL AS WELL AS  $\mu_E$  ON MESSIDOR DATA SET

	$E \leq 0.05$	$E \leq 0.1$	$E \leq 0.15$	$E \leq 0.2$	$E \leq 0.25$	$\mu_E$
CHT [27]	7%	46%	73%	84%	90%	14.0%
Proposed	8%	51%	76%	86%	92%	12.5%

TABLE V  
MEAN OVERLAPPING ERROR  $\mu_E$  FROM TESTING IMAGES AT DIFFERENT PARAMETERS FOR CUP SEGMENTATION

	no filter	$5 \times 5$	$10 \times 10$	$15 \times 15$	$20 \times 20$	$25 \times 25$
50	27.4%	26.4%	25.7%	25.4%	25.0%	25.1%
100	25.0%	24.4%	24.2%	24.1%	24.2%	24.3%
200	26.0%	25.6%	25.4%	25.2%	25.3%	25.3%

TABLE VI  
PERCENTAGE OF IMAGES PER  $E$  INTERVAL AND  $\mu_E$  FOR CUP SEGMENTATION

	$E < 0.1$	$E < 0.2$	$E < 0.3$	$E < 0.4$	$E < 0.5$	$\mu_E$
$HIST_j^c$	1%	4%	13%	24%	43%	53.6%
$HIST_j^c + \widehat{CSS}_j^c$	5%	35%	67%	87%	94%	27.0%
$HIST_j^c + \widehat{CSS}_j^c + D_j$	8%	42%	75%	90%	96%	24.1%

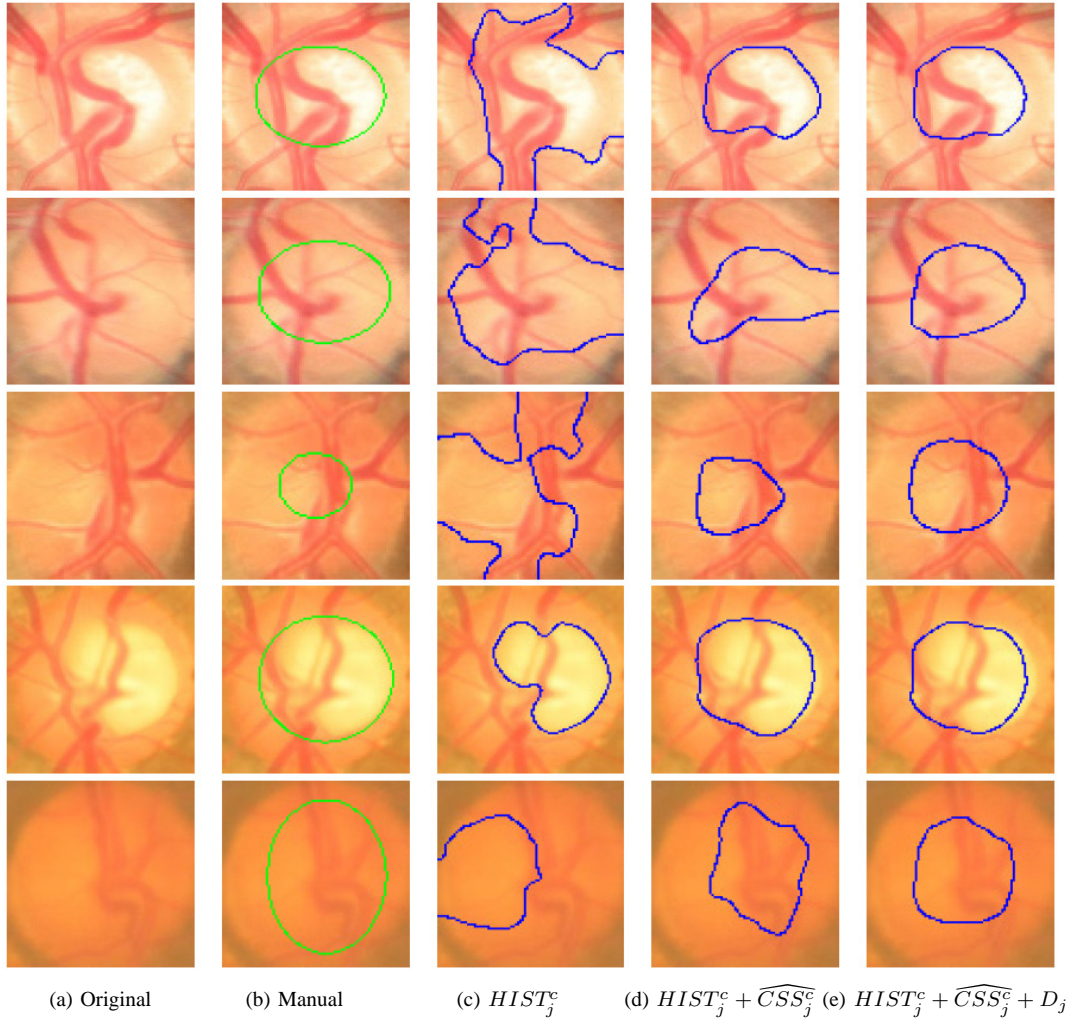


Fig. 12. Sample results. From left to right columns: (a) the original images, (b) the manual “ground truth”, (c)-(e) outlines by the proposed method before ellipse fitting using  $HIST_j^c$  only,  $HIST_j^c + \widehat{CSS}_j^c$  and  $HIST_j^c + \widehat{CSS}_j^c + D_j$ .



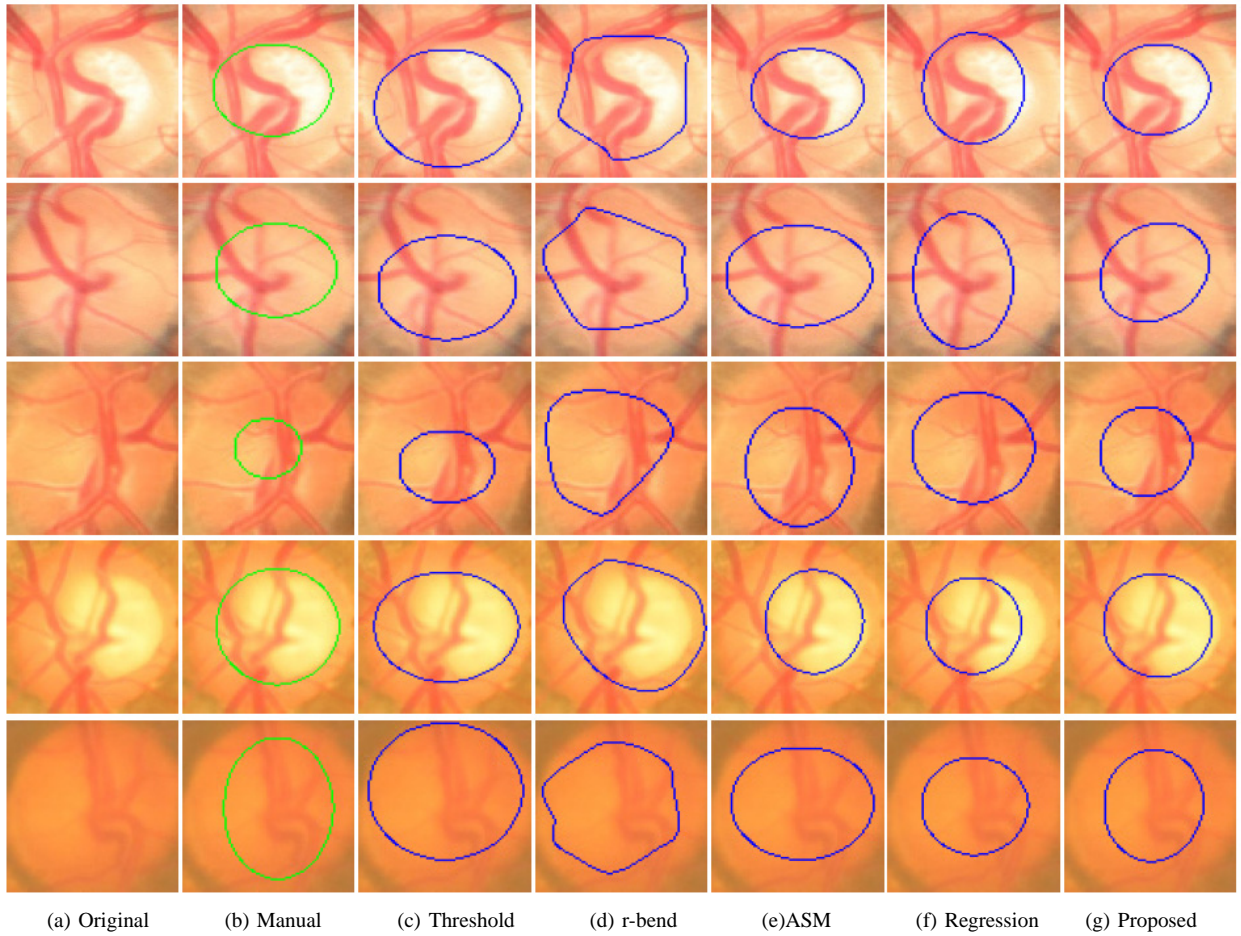


Fig. 13. Sample results. From left to right columns: (a) the original images, (b) the manual “ground truth”, (c)-(f) outlines by threshold [56], r-bend [8], ASM [59], regression [58] and the proposed method. From top to bottom, the manual CDR for the five examples from top to bottom are: 0.54, 0.55, 0.35, 0.68 and 0.83. The CDR by the proposed method are 0.53, 0.58, 0.52, 0.61, and 0.66 respectively.

TABLE VII  
PERCENTAGE OF IMAGES PER  $E$  INTERVAL AND  $\mu_E$  FOR CUP SEGMENTATION

	$E \leq 0.1$	$E \leq 0.2$	$E \leq 0.3$	$E \leq 0.4$	$E \leq 0.5$	$\mu_E$
Thresholding [56]	0%	3%	15%	31%	47%	53.5%
R-bend [8]	0%	4%	28%	56%	77%	39.5%
ASM [31]	3%	25%	51%	76%	88%	31.3%
Regression [58]	6%	29%	62%	81%	95%	28.4%
Proposed	8%	42%	75%	90%	96%	24.1%

#### D. Glaucoma Screening

This paper also evaluates the accuracy of the proposed CDR and its performance in glaucoma screening. The CDR error is computed as

$$\delta = |CDR_{GT} - CDR|, \quad (8)$$

where  $CDR_{GT}$  denotes the manual CDR from trained professionals. We now evaluate the proposed method using the 325 SiMES testing images and 1676 SCES images. Proposed disc segmentation is used to get the discs.

We compare the proposed cup segmentation method with previous cup segmentation methods including the threshold method [56], r-bend method [8], ASM method [59] and regression method [58]. Fig. 14 shows the ROC curves of

these methods and Table VIII summarizes the mean CDR error  $\mu_\delta$ , the AUC of the ROC curves by various cup segmentation methods. In addition, the  $p$ -values of the ROC comparisons between proposed method and other methods are computed to show the significance of the improvement. The ROC comparison is done using the approach proposed by Delong *et al.* [64]. According to Delong’s method, a  $p$ -value smaller than 0.05 indicates a significant improvement. Therefore, the proposed method achieves AUC significantly larger than IOP, threshold, r-bend, ASM, and regression methods. The results show smaller CDR errors in CDR measurement and higher AUC in glaucoma screening by the proposed method, compared with previous methods. The proposed disc and cup segmentation methods achieve an AUC of 0.800, 0.039 lower



TABLE VIII

MEAN CDR ERROR  $\mu_\delta$ , AUC BY VARIOUS METHODS AND THE  $p$ -VALUES OF ROC COMPARISON BETWEEN PROPOSED METHOD AND OTHER METHODS.

Measurement	Mean CDR error $\mu_\delta$		ROC Comparison			
Data set	SiMES		SiMES		SCES	
	glaucoma	healthy	AUC	$p$ -value (vs. Proposed)	AUC	$p$ -value (vs. Proposed)
Airpuff IOP	-	-	0.585	$p < 0.0001$	0.660	$p = 0.0039$
Threshold [56]	0.142	0.131	0.634	$p < 0.0001$	0.574	$p < 0.0001$
R-bend [8]	0.107	0.179	0.653	$p < 0.0001$	0.615	$p < 0.0001$
ASM [59]	0.130	0.094	0.707	$p < 0.0001$	0.756	$p = 0.0389$
Regression [58]	0.138	0.093	0.715	$p = 0.0003$	0.767	$p = 0.0394$
Proposed	0.107	0.077	0.800	-	0.822	-
Manual	-	-	0.839	$p = 0.4949$	-	-

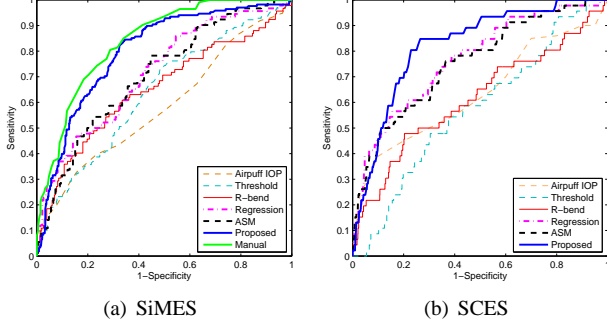


Fig. 14. ROC curves of glaucoma screening by previous and proposed cup segmentation methods in SiMES and SCES.

than AUC of 0.839 of the manual CDR computed from manual disc and manual cup. In the results for the SCES data set, the proposed method achieves AUC 0.822 in the screening SCES data, which is much higher than 0.660 by the currently used IOP measurement. From the discussions with clinicians, the accuracy is good enough for a large-scale glaucoma screening program.

Since there is many possibilities to divide the 650 images for training and testing, it is important to know how different partition affects the performance. In our experiments, we randomly obtain 10 different random partitions and conduct two-fold cross validation in each partition. Each partition divides the images to set  $A$  and  $B$  with 325 images. In the cross-validation, set  $A$  is first used to train new models for disc and cup segmentation and set  $B$  is used for testing. Then set  $B$  is used for training and set  $A$  is used for testing. For SCES data set, it remains the same in each test and only the models are updated each time. The AUC in each test is computed. Therefore we get 20 different results for both SiMES and SCES from the 10 partitions. We utilize the MATLAB (MathWorks, Inc.) box plot to describe the results. Each box plot produces a box and whisker plot for each data set in Fig. 15. The boxes have lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the boxes to show the extent of the rest of the data. Maximum whisker length is 1.5 times the interquartile range.

## V. DISCUSSIONS AND CONCLUSIONS

We have presented superpixel classification based methods for disc and cup segmentations for glaucoma screening. It

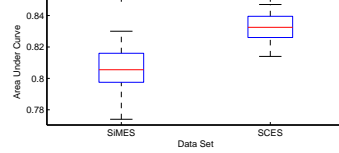


Fig. 15. Boxplot of AUC in glaucoma screening from ten times two-fold cross validation

has been demonstrated that CSS is beneficial for both disc and cup segmentation. In disc segmentation, HIST and CSS complement each other. CSS responds to blobs and provides better differentiation between PPA and discs compared with histograms. Histograms with the contrast enhancement overcome the limitation of CSS due to contrast variations. Reliability score is an important indicator of the automated results. From our experience, disc segmentations with  $r \geq 0.85$  is likely to indicate good cases. For lower ones, it is likely that the results are inaccurate, even though the deformation in the last step might still find a good result in some situations. It is important to have a good disc segmentation because the CDR computed from a wrong disc is not very meaningful for doctors. In cup segmentation, the benefit of CSS is even larger than that in disc segmentation, because the colour change from cup to neuroretinal rim is much smaller. Therefore, the uneven illumination becomes a large noise affecting the cup segmentation. The CSS computed from the centre surround difference is less sensitive, and thereby improves the result. It is important to point out that the proposed superpixel classification is used as an initialization for deformable models. We have demonstrated that, by replacing circular Hough transform based initialization with the proposed one for active shape model, we are able to improve the disc segmentation.

One limitation of the proposed cup segmentation is that the trained classifier is slightly dominated by cups with medium sizes, so the proposed method under-estimates the very large cups, while over-estimating the very small cups when pallor is not obvious. This is partly due to the use of the location feature  $D_j$ . However, the advantage is that  $D_j$  helps to avoid very bad results. Both the cup segmentation accuracy and glaucoma screening accuracy show improvement by including  $D_j$  in the feature space. One possible solution to overcome this limitation is to collect more samples with very small and very large cups for training. Another possible solution is to adopt multiple kernels. This paper concatenates the three types of

features from different sources. In future work, multiple kernel learning [65] will be used for enhancement. A third possible solution is to use vessel bends to correct the bias in current cup segmentation. For example, we can obtain the vessel bends as in [57][8] from the vicinity of the current cup boundary to fine-tune the cup boundary. As it requires some heuristics in vessel tracking and bending detection, a robust algorithm is important for future development.

Despite the bias, the obtained CDRs for very large cups are still high and the CDRs for very small cups are still small, so the proposed method achieves high classification accuracy when differentiating glaucomatous subjects from healthy ones. The accuracy of the proposed method is already much better than the airpuff IOP measurement and previous CDR based methods [56][58][59]. The glaucoma detection accuracy in SiMES by the proposed method is close to the manual measurement. The results by the proposed methods are good enough for a screening purposes in polyclinics and eye centres, according to discussions with clinicians and ophthalmologists. However, there are still many features that can be improved in the proposed method, including the bias we discussed. The proposed methods in this paper model the disc/cup boundary as an ellipse, which is an approximation of the actual boundary. The modelling is beneficial as the disc/cup boundary can be unclear at some sectors. For example, the inferior and superior of the disc boundaries are often affected by the blood vessels entering the disc. The cup boundary at the nasal side of the cup is often difficult to be determined even manually due to the presence of blood vessels. The modelling approach helps to identify the disc and cup boundary at these sectors. The disadvantage is that the modelling may not cover some morphology such as a sudden change of disc or cup boundary. The CDR based screening from 2D images also has its limitations. The disc is a 3D structures with depth the primary indicator for cup. However, 2D images do not have depth information. Compared with 3D images which capture true 3D morphological structures of disc and cup, 2D images capture the color information of disc and rely on color change to get the cup boundary. A weak color change would increase the challenge in cup segmentation. This paper utilized the prior knowledge of cup location by using location feature to overcome this problem. Although some improvement can be seen, it also leads to the systematic bias. Since CDR only reflects one aspect of the disc, combining it with other factors is expected to further improve the performance. Future work will explore the integration of other factors to improve diagnostic outcomes towards a more reliable and efficient glaucoma screening system.

#### ACKNOWLEDGMENT

The authors would like to thank the MESSIDOR program partners (see <http://messidor.crihan.fr>) and Analysis of Digital Retinal Images Project team (see <http://www.uhu.es/retinopathy>) for providing the MESSIDOR database and the manual optic disc ground truth.

#### REFERENCES

- [1] H. A. Quigley and A. T. Broman, "The number of people with glaucoma worldwide in 2010 and 2020," *Br. J. Ophthalmol.*, vol. 90(3), pp. 262–267, 2006.
- [2] S. Y. Shen, T. Y. Wong, P. J. Foster, J. L. Loo, M. Rosman, S. C. Loon, W. L. Wong, S. M. Saw, and T. Aung, "The prevalence and types of glaucoma in malay people: the singapore malay eye study," *Invest. Ophthalmol. Vis. Sci.*, vol. 49(9), pp. 3846–3851, 2008.
- [3] P. J. Foster, F. T. Oen, D. Machin, T. P. Ng, J. G. Devereux, G. J. Johnson, P. T. Khaw, and S. K. Seah, "The prevalence of glaucoma in chinese residents of singapore: a cross-sectional population survey of the tanjong pagar district," *Arch. Ophthalmol.*, vol. 118(8), pp. 1105–1111, 2000.
- [4] Centre for Eye Research Australia, *Tunnel vision : the economic impact of primary open angle glaucoma. [electronic resource]*, 2008, <http://nla.gov.au/nla.arc-86954>.
- [5] J. Meier, R. Bock, G. Michelson, L. G. Nyl, and J. Hornegger, "Effects of preprocessing eye fundus images on appearance based glaucoma classification," *Proc. Computer Analysis of Images and Patterns*, vol. 4673, pp. 165–172, 2007.
- [6] R. Bock, J. Meier, G. Michelson, L. G. Nyl, and J. Hornegger, "Classifying glaucoma with image-based features from fundus photographs," *Proc. of DAGM*, pp. 355–364, 2007.
- [7] R. Bock, J. Meier, L. G. Nyl, and G. Michelson, "Glaucoma risk index: Automated glaucoma detection from color fundus images," *Med. Image Anal.*, vol. 14, pp. 471–481, 2010.
- [8] G. D. Joshi, J. Sivaswamy, and S. R. Krishnadas, "Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment," *IEEE Trans. Med. Imag.*, vol. 30, pp. 1192–1205, 2011.
- [9] T. Damms and F. Dannheim, "Sensitivity and specificity of optic disc parameters in chronic glaucoma," *Invest. Ophth. Vis. Sci.*, vol. 34, pp. 2246–2250, 1993.
- [10] D. Michael and O. D. Hancox, "Optic disc size, an important consideration in the glaucoma evaluation," *Clinical Eye and Vision Care*, vol. 11, pp. 59–62, 1999.
- [11] N. Harizman, C. Oliveira, A. Chiang, C. Tello, M. Marmor, R. Ritch, and J. M. Liebmann, "The isnt rule and differentiation of normal from glaucomatous eyes," *Arch. Ophthalmol.*, vol. 124, pp. 1579–1583, 2006.
- [12] J. B. Jonas, M. C. Fernandez, and G. O. Naumann, "Glaucomatous parapapillary atrophy occurrence and correlations," *Arch Ophthalmol*, vol. 110, pp. 214–222, 1992.
- [13] R. R. Allingham, K. F. Damji, S. Freedman, S. E. Moroi, and G. Shafra-nov, *Shields' Textbook of Glaucoma, 5th Edition*, Lippincott Williams & Wilkins, 2005.
- [14] J. Xu, O. Chutatape, E. Sung, C. Zheng, and P. C. T. Kuan, "Optic disk feature extraction via modified deformable model technique for glaucoma analysis," *Pattern Recognition*, vol. 40, pp. 2063–2076, 2007.
- [15] M. D. Abràmoff, W. L. M. Alward, E. C. Greenlee, L. Shuba, C. Y. Kim, J. H. Fingert, and Y. H. Kwon, "Automated segmentation of the optic disc from stereo color photographs using physiologically plausible features," *Invest. Ophthalmol. Vis. Sci.*, vol. 48, pp. 1665–1673, 2007.
- [16] Z. Hu, M. D. Abràmoff, Y. H. Kwon, K. Lee, and M. K. Garvin, "Automated segmentation of neural canal opening and optic cup in 3-d spectral optical coherence tomography volumes of the optic nerve head," *Inv Ophthalmol Vis Sci.*, vol. 51, pp. 5708–5717, 2010.
- [17] M. D. Abràmoff, K. Lee, M. Niemeijer, W. L. M. Alward, E. Greenlee, M. Garvin, M. Sonka, and Y. H. Kwon, "Automated segmentation of the cup and rim from spectral domain oct of the optic nerve head," *Inv Ophthalmol Vis Sci.*, vol. 50, pp. 5778–5784, 2009.
- [18] K. Stapor, A. Sacutewitonski, R. Chrastek, and G. Michelson, "Segmentation of fundus eye images using methods of mathematical morphology for glaucoma diagnosis," in *Proc. Int. Conf. Computer Science*, pp. 41–48, 2004.
- [19] N. Inoue, K. Yanashima, K. Magatani, and T. Kurihara, "Development of a simple diagnostic method for the glaucoma using ocular fundus pictures," *Int. Conf. of IEEE Eng. in Med. and Bio. Soc.*, pp. 3355–3358, 2005.
- [20] D. W. K. Wong, J. Liu, J. H. Lim, X. Jia, H. Li F. Yin and, and T. Y. Wong, "Level-set based automatic cup-to-disc ratio determination using retinal fundus images in argali," *Int. Conf. of IEEE Eng. in Med. and Bio. Soc.*, pp. 2266–2269, 2008.
- [21] Y. Hatanaka, A. Noudo, C. Muramatsu, A. Sawada, T. Hara, T. Yamamoto, and H. Fujita, "Automatic measurement of vertical cup-to disc ratio on retinal fundus images," in *Proc. Int. Conf. Medical Biometrics*, pp. 64–72, 2010.

- [22] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imaging*, vol. 23, no. 4, pp. 501–509, 2004.
- [23] A. Hoover and M. Goldbaum, "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels," *IEEE Trans. Med. Imag.*, vol. 22, pp. 951–958, 2003.
- [24] M. Foracchia, E. Grisan, and A. Ruggeri, "Detection of optic disc in retinal images by means of a geometrical model of vessel structure," *IEEE Trans. Med. Imag.*, vol. 23, no. 10, pp. 1189–1195, 2004.
- [25] Z. Zhang, B. H. Lee, J. Liu, D. W. K. Wong, N. M. TAN, J. H. Lim, F. S. Yin, W. M. Huang, and H. Li, "Optic disc region of interest localization in fundus image for glaucoma detection in argali," *Proc. of Int. Conf. on Industrial Electronics & Applications*, pp. 1686–1689, 2010.
- [26] X. Zhu and R. M. Rangayyan, "Detection of the optic disc in images of the retina using the hough transform," *Int. Conf. of IEEE Eng. in Med. and Bio. Soc.*, pp. 3546–3549, 2008.
- [27] A. Aquino, M. Gegundez-Arias, and D. Marin, "Detecting the optic disc boundary in digital fundus images using morphological, edge detection, and feature extraction techniques," *IEEE Trans. Med. Imag.*, vol. 29, pp. 1860–1869, 2010.
- [28] J. Cheng, J. Liu, D. W. K. Wong, F. Yin, C. Cheung, M. Baskaran, T. Aung, and T. Y. Wong, "Automatic optic disc segmentation with peripapillary atrophy elimination," *Int. Conf. of IEEE Eng. in Med. and Bio. Soc.*, pp. 6624–6627, 2011.
- [29] J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, E. Fletcher, and L. Kennedy, "Optic nerve head segmentation," *IEEE Trans. Med. Imag.*, vol. 23, no. 2, pp. 256–264, 2004.
- [30] Z. Zhang, J. Liu, N. S. Cherian, Y. Sun, J. H. Lim, W. K. Wong, N. M. Tan, S. Lu, H. Li, and T. Y. Wong, "Convex hull based neuro-retinal optic cup ellipse optimization in glaucoma diagnosis," *Int. Conf. of IEEE Eng. in Med. and Bio. Soc.*, pp. 1441–1444, 2009.
- [31] F. Yin, J. Liu, S. H. Ong, Y. Sun, D. W. K. Wong, N. M. Tan, C. Cheung, M. Baskaran, T. Aung, and T. Y. Wong, "Model-based optic nerve head segmentation on retinal fundus images," *Int. Conf. of IEEE Eng. in Med. and Bio. Soc.*, pp. 2626–2629, 2011.
- [32] L. Tang, M. K. Garvin, Y. H. Kwon, and M. D. Abramoff, "Segmentation of optic nerve head rim in color fundus photographs by probability based active shape model," *ARVO*, 2012.
- [33] D. W. K. Wong, J. Liu, N. M. Tan, F. Yin, B. H. Lee, and T. Y. Wong, "Learning-based approach for the automatic detection of the optic disc in digital retinal fundus photographs," *Int. Conf. of IEEE Eng. in Med. and Bio. Soc.*, pp. 5355–5358, 2010.
- [34] J. B. Jonas, W. M. Budde, and S. Panda-Jonas, "Ophthalmoscopic evaluation of the optic nerve head," *Surv. Ophthalmol.*, pp. 293–320, 1999.
- [35] J. Cheng, J. Liu, Y. Xu, D. W. K. Wong, B. H. Lee, C. Cheung, T. Aung, and T. Y. Wong, "Superpixel classification for initialization in model based optic disc segmentation," *Int. Conf. of IEEE Eng. in Med. and Bio. Soc.*, pp. 1450–1453, 2012.
- [36] J. A. Giacony, S. K. Law, A. L. Coleman, and J. Caprioli, *Pearls of Glaucoma Management*, Springer Heidelberg Dordrecht London New York, 2010.
- [37] C. Muramatsu, T. Nakagawa, A. Sawada, Y. Hatanaka, T. Hara, T. Yamamoto, and H. Fujita, "Automated segmentation of optic disc region on retinal fundus photographs: Comparison of contour modeling and pixel classification methods," *Computer Methods and Programs in Biomedicine*, vol. 101, pp. 23–32, 2011.
- [38] X. Ren and J. Malik, "Learning a classification model for segmentation," *Int. Conf. Computer Vision*, vol. 1, pp. 10–17, 2003.
- [39] P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," *Int'l J. Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.
- [40] T. Cour, F. Benezit, and J. Shi, "Spectral segmentation with multiscale graph decomposition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2005.
- [41] A. Moore, S. Prince, J. Warrell, U. Mohammed, and G. Jones, "Superpixel lattices," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2008.
- [42] O. Veksler, Y. Boykov, and P. Mehrani, "Superpixels and supervoxels in an energy optimization framework," *Proc. European Conf. Computer Vision*, vol. V, pp. 211–224, 2010.
- [43] A. Vedaldi and S. Soatto, "Quick shift and kernel methods for mode seeking," *Proc. European Conf. Computer Vision*, vol. IV, pp. 705–718, 2008.
- [44] A. Levinshtein, A. Stere, K. Kutulakos, D. Fleet, S. Dickinson, and K. Siddiqi, "Turbopixels: Fast superpixels using geometric flows," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 31, no. 12, pp. 2290–2297, 2009.
- [45] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 24, no. 5, pp. 603–619, 2002.
- [46] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 13, no. 6, pp. 583–598, 1991.
- [47] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 34, pp. 2274–2281, 2012.
- [48] J. Tighe and S. Lazebnik, "Superparsing: Scalable nonparametric image parsing with superpixels," *European Conf. on Computer Vision*, vol. 5, pp. 352–365, 2010.
- [49] T. H. Hildebrandt, "A local neural implementation of histogram equalization," *IEEE Int. Conf. on Neural Networks*, vol. 3, pp. 1678–1683, 1993.
- [50] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden, "Pyramid methods in image processing," *RCA Engineer*, vol. 29(6), pp. 33–41, 1984.
- [51] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, 1998.
- [52] D. Song and D. Tao, "Biologically inspired feature manifold for scene classification," *IEEE Trans. Image Processing*, vol. 19, no. 1, pp. 174–184, 2010.
- [53] J. Cheng, D. Tao, J. Liu, D. W. K. Wong, N. M. Tan, T. Y. Wong TY, and S. M. Saw, "Peripapillary atrophy detection by sparse biologically inspired feature manifold," *IEEE Trans. Med. Imaging*, vol. 31, pp. 2355–2365, 2012.
- [54] C. C. Chang and C. J. Lin, *LIBSVM: a library for support vector machines*, 2001, Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [55] Z. Kalal, J. Matas, and K. Mikolajczyk, "Weighted Sampling for Large-Scale Boosting," *British Machine Vision Conference*, 2008.
- [56] G. D. Joshi, J. Sivaswamy, K. Karan, and R. Krishnadas, "Optic disk and cup boundary detection using regional information," in *Proc. IEEE Int. Symp. Biomed. Imag.*, pp. 948–951, 2010.
- [57] D. W. K. Wong, J. Liu, J. H. Lim, H. Li, and T. Y. Wong, "Automated detection of kinks from blood vessels for optic cup segmentation in retinal images," *Proc. SPIE 7260*, p. 72601J, 2009.
- [58] Y. Xu, D. Xu, S. Lin, J. Liu, J. Cheng, C. Y. Cheung, T. Aung, and T. Y. Wong, "Sliding window and regression based cup detection in digital fundus images for glaucoma diagnosis," *MICCAI 2011, Part III. LNCS*, pp. 1–8, 2011.
- [59] F. Yin, J. Liu, D. W. K. Wong, N. M. Tan, C. Cheung, M. Baskaran, T. Aung, and T. Y. Wong, "Automated segmentation of optic disc and optic cup in fundus images for glaucoma diagnosis," *IEEE Int. Symp. on Computer-Based Medical Systems*, pp. 1–6, 2012.
- [60] A. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least squares fitting of ellipses," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 21(5), pp. 476–480, 1999.
- [61] Z. Zhang, F. Yin, J. Liu, W. K. Wong, N. M. Tan, B. H. Lee, J. Cheng, and T. Y. Wong, "Origami-light: An online retinal fundus image database for glaucoma analysis and research," *Int. Conf. of IEEE Eng. in Med. and Bio. Soc.*, pp. 3065–3068, 2010.
- [62] MESSIDOR TECHNO-VISION Project France Download Images Section, MESSIDOR: Digital Retinal Images, [Online]. Available: <http://messidor.crihan.fr/download-en.php>.
- [63] Spain Univ. Huelva [Online]. Expert system for early automated detection of DR by analysis of digital retinal images project website, Huelva, Available: <http://www.uhu.es/retinopathy>.
- [64] E. R. Delong, D. M. Delong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach," *Biometrics*, vol. 44, no. 3, pp. 837–845, 1988.
- [65] L. Duan, I. W. Tsang, and D. Xu, "Domain transfer multiple kernel learning," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 34, pp. 465–479, 2012.